CUSTOMER SHOPPING BEHAVIOR ANALYSIS

1. Project Introduction

This project focuses on examining customer shopping patterns by analyzing transactional data collected from 3,900 individual purchases spanning multiple product categories. The primary objective is to extract meaningful insights related to customer spending habits, purchasing behavior, product preferences, and subscription usage, enabling data-driven business decisions.

2. Dataset Overview

The analysis is based on a structured dataset containing both customer-level and transaction-level information.

- **Total Records:** 3,900 transactions
- **Total Attributes:** 18 variables

Key Data Attributes

- **Customer Information:** Age, Gender, Location, Subscription Status
- **Transaction Details:** Item Purchased, Product Category, Purchase Amount, Season, Size, Color
- **Behavioral Metrics:** Discount Applied, Promo Code Usage, Previous Purchases, Purchase Frequency, Review Ratings, Shipping Type

Data Quality Notes

- The *Review Rating* column contained 37 missing values, which were addressed during preprocessing.

3. Exploratory Data Analysis (Python)

Python was used for initial data exploration, cleaning, and feature preparation.

Data Preparation Steps

- **Data Import:** Loaded the dataset using the pandas library.
- **Structural Review:** Assessed data types and structure using `.info()` and statistical summaries via `.describe()`.

```
df.describe(include='all')
```

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900	3900.000000	3900	3900
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	2	2	NaN	6	7
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	No	No	NaN	PayPal	Every 3 Months
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	2223	2223	NaN	677	584
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN	NaN	25.351538	NaN	NaN
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN	NaN	14.447125	NaN	NaN
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	NaN	NaN	1.000000	NaN	NaN
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	NaN	NaN	13.000000	NaN	NaN
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	NaN	NaN	25.000000	NaN	NaN
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	NaN	NaN	38.000000	NaN	NaN
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	NaN	NaN	50.000000	NaN	NaN

- **Handling Missing Values:** Imputed missing review ratings using the median rating within each product category to maintain consistency.
- **Column Formatting:** Standardized column names into snake_case format for better readability and documentation.

Feature Engineering

- Generated an **age_group** variable by categorizing customers into defined age ranges.
- Created **purchase_frequency_days** to better represent customer buying intervals.

Data Validation & Optimization

- Identified redundancy between *discount_applied* and *promo_code_used* variables; removed the promo code field.
- Loaded the cleaned dataset into a PostgreSQL database to support SQL-based analysis.

4. SQL-Based Business Analysis

Using MYSQL, multiple business-focused queries were executed to uncover insights.

Key Analytical Questions Addressed

1. **Revenue by Gender:** Compared total revenue contributions across male and female customers.

gender	revenue
Male	157890
Female	75191

2. **High-Value Discount Shoppers:** Identified customers who utilized discounts yet spent more than the average purchase amount.

customer_id	purchase_amount
2	64
3	73
4	90
7	85
9	97
12	68
13	72
16	81
20	90
22	62
24	88
29	94
32	79
33	67
35	91
37	69

3. **Top 5 Products by Rating :** Determined the top five products based on average customer review ratings.

item_purchased	avg product rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.8
Skirt	3.78

4. **Shipping Method Impact:** Analyzed differences in average purchase value between standard and express shipping.

shipping_type	round(avg(purchase_amount),2)
Express	60.48
Standard	58.46

5. **Subscriber vs. Non-Subscriber Performance:** Compared average spending and total revenue across subscription statuses.

subscription_status	total customers	average spend	total revenue
Yes	1053	59.49	1634061
No	2847	59.87	4285659

6. **Discount-Sensitive Products:** Identified five products with the highest proportion of discounted purchases.

item_purchased	discount rate
Sneakers	50
Hat	50
Coat	49
Sweater	48

7. **Customer Segmentation:** Categorized customers into New, Returning, and Loyal groups based on purchase history.

customer_segment	number of customer
loyal	3116
returning	701
new	83

8. **Top Products by Category:** Listed the three most frequently purchased products within each category.

item_rank	category	item_purchased	total_orders
1	Accessories	Jewelry	171
2	Accessories	Sunglasses	161
3	Accessories	Belt	161
1	Clothing	Blouse	171
2	Clothing	Pants	171
3	Clothing	Shirt	169
1	Footwear	Sandals	160
2	Footwear	Shoes	150
3	Footwear	Sneakers	145
1	Outerwear	Jacket	163
2	Outerwear	Coat	161

9. **Repeat Purchase & Subscription Link:** Evaluated whether customers with more than five purchases showed higher subscription adoption.

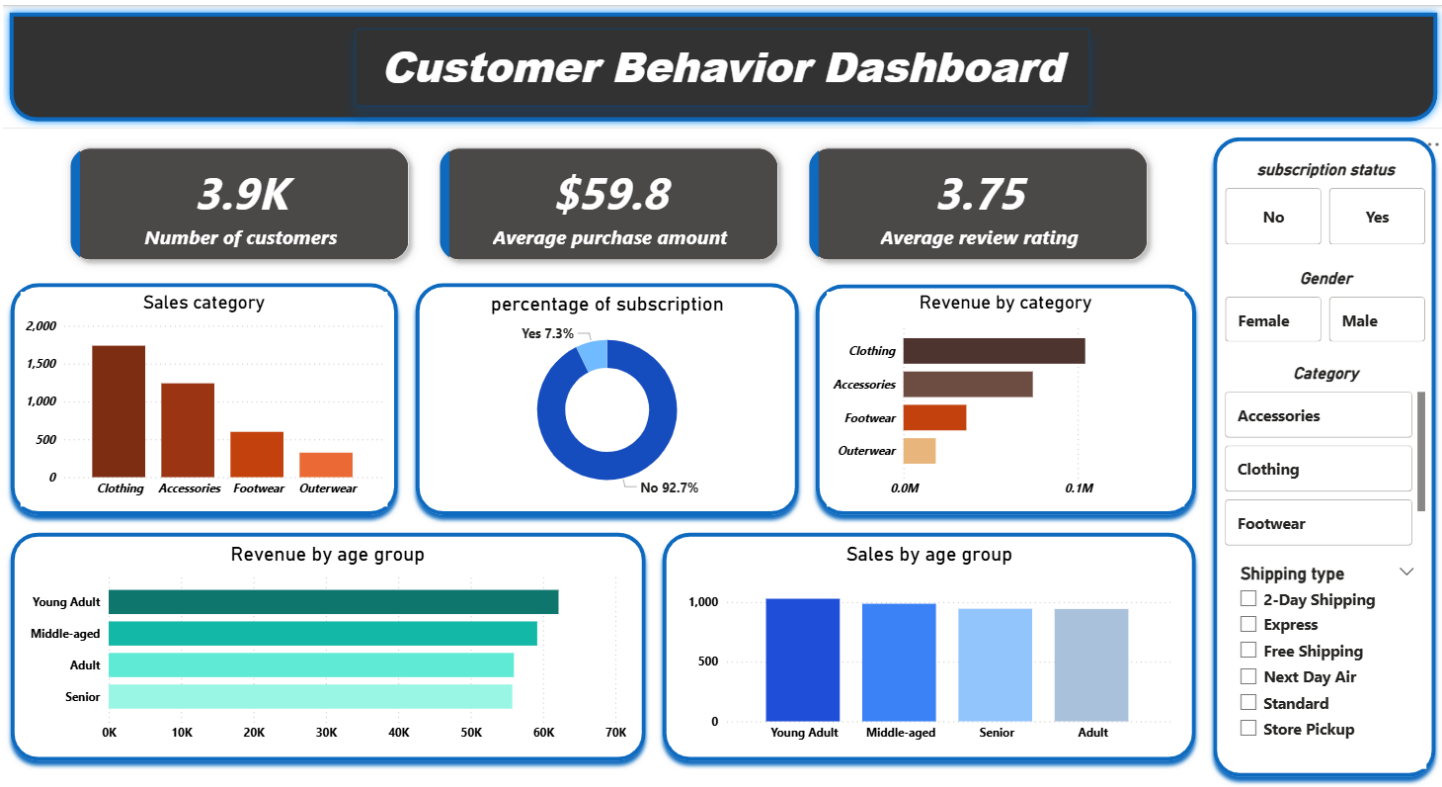
subscription_status	repeate_buyers
Yes	958
No	2518

10. **Revenue Distribution by Age Group:** Calculated revenue contributions across different age segments.

age_group	total_revenue
Young Adult	62143
Middle-aged	59197
Adult	55978
Senior	55763

5. Power BI Dashboard

An interactive Power BI dashboard was developed to visually communicate insights and trends. The dashboard highlights: - Total number of customers - Average purchase amount - Average review rating - Revenue and sales breakdown by category - Subscription distribution - Revenue contribution by age group



6. Business Insights & Recommendations

Based on the analysis, the following strategic recommendations are proposed:

- **Increase Subscription Adoption:** Promote exclusive offers and benefits to encourage non-subscribers to enroll.
- **Strengthen Customer Loyalty Programs:** Incentivize repeat purchases to convert returning customers into loyal ones.
- **Optimize Discount Strategies:** Balance discount usage to drive sales while protecting profit margins.
- **Enhance Product Promotion:** Focus marketing campaigns on top-rated and best-selling products.
- **Targeted Marketing Efforts:** Allocate resources toward high-revenue age groups and customers preferring express shipping.
